UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/684,488	10/04/2000	Bin Zhang	10992482-1	3131

7590 07/31/2003

HEWLETT-PACKARD COMPANY
Intellectual Property Administration
P.O. Box 272400
Fort Collins, CO 80528-9599

EXAMINER
HAMILTON, MONPLAISIR G

ART UNIT	PAPER NUMBER
2172	

DATE MAILED: 07/31/2003

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)
	09/684,488	ZHANG ET AL.
	Examiner	Art Unit
	Monplaisir G Hamilton	2172

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *17 June 2003* .

2a)☐ This action is **FINAL**.   2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-10* is/are pending in the application.

    4a) Of the above claim(s) *11-20* is/are cancelled.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-10* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

11)☐ The proposed drawing correction filed on _____ is: a)☐ approved b)☐ disapproved by the Examiner.

    If approved, corrected drawings are required in reply to this Office action.

12)☐ The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. §§ 119 and 120

13)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____ .

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

14)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).

    a) ☐ The translation of the foreign language provisional application has been received.

15)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

Attachment(s)

1)☒ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☐ Information Disclosure Statement(s) (PTO-1449) Paper No(s) _____ .

4)☐ Interview Summary (PTO-413) Paper No(s). _____ .

5)☐ Notice of Informal Patent Application (PTO-152)

6)☐ Other: .

DETAILED ACTION

Continued Examination Under 37 CFR 1.114

1.   A request for continued examination under 37 CFR 1.114, including the fee set forth in

37 CFR 1.17(e), was filed in this application after final rejection. Since this application is

eligible for continued examination under 37 CFR 1.114, and the fee set forth in 37 CFR 1.17(e)

has been timely paid, the finality of the previous Office action has been withdrawn pursuant to

37 CFR 1.114. Applicant's submission filed on 6/17/03 has been entered.

Response to Arguments

2.   Applicant's arguments filed 6/17/03 have been fully considered but they are not

persuasive.

Applicant's argue "..Thus, the parameter c in the Guha reference is not used to determine

how many points should be moved from cluster to cluster, rather, the parameter c is used only to

determine how many of the pints should be used to calculate a representative point for distance

calculation purposes. Thus Guha does not teach "the C parameter specifies the number of data

points that will be evaluated when deciding to merge a pair of clusters...". Further, Guha does

not teach the "size parameter (C) specifies the number of data points to be moved at one time

from one cluster to another."

Examiner contends that Guha's disclosure of the C parameter is equivalent to claimed

sized parameter for the following reasons. Guha, col 10, lines 35-50 discloses the merged

procedure chooses c scattered points from among all the points in the merged cluster, instead, if

we selected the c points of w from the clusters u an v, which are merged to from w, 2C points

instead of O(n) points need to be examined when selecting representative points, reducing the

complexity O(1). Examiner holds that this disclosure teaches that C is used to determine which

points are selected (moved) to represent a merged cluster. Examiner has interpreted the above

choosing or selecting to be equivalent to the claimed moving. Therefore, the C parameter

represents the number of points to be moved from one cluster to another.

Claim Rejections - 35 USC § 112

3.   Claims 1-10 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter, which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention. On, page 16, lines 10-14 applicant has set forth applications or programs that can use the claimed method, however examiner maintains that this disclosure does not sufficiently disclose the claimed inter-relationship between **hardware** and software that is necessary to enable a computer system. Applicant has argued that one of ordinary skill after reading the aforementioned section of specification would understand that the method is implemented in a computer system. Examiner agrees that one of ordinary skill can understand that a computer is being used however there is no disclosure about the type of system that will be used to implement the claimed invention, is the system is a single processor, multiple processor system, is it a mainframe system, examiner holds that without a proper disclosure of the type of system being used, one of ordinary skill would not be able to build the claimed system. Examiner therefore concludes that the claimed microprocessor/computer system is not disclosed anywhere in the specification. Therefore this amendment to claim adds new matter.

Claims 1-10 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the enablement requirement. The claim(s) contains subject matter, which was not described in the specification in such a way as to enable one skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention. Examiner holds that the claimed computer-readable medium and microprocessor are not described anywhere in the

specification. Examiner holds that the disclosure on page 16, lines 10 and 14 does not teach the

type of hardware required to enable the claimed computer system. Furthermore as argued by

applicant Fig. 2 and 3 discloses software modules. These modules do not show an inter-

relationship with the newly claimed hardware. The drawings and the specification do not

sufficiently support the newly claimed computer-readable medium or microprocessor.

Therefore, the newly added limitations are not enabled by the specification.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in
> section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are
> such that the subject matter as a whole would have been obvious at the time the invention was made to a person
> having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the
> manner in which the invention was made.

4.   Claims 1-10 are rejected under 35 U.S.C. 103(a) as being unpatentable over US Patent

issued to Guha et al, herein referred to as Guha.

Referring to Claim 1:

Guha discloses receiving a plurality of data points for clustering (col 5, lines 60-65); (b)

receiving a size parameter for specifying the number of data points to be moved at one time (col

6, lines 50-61); (c) clustering the data points by using the size parameter to generate clustered

results (col 6, lines 55-60) (d) determining whether the clustered results are satisfactory (col 4,

lines 40-45); (e) when the clustered results are satisfactory, stop clustering (col 4, lines 40-45)."

Guha further discloses a method that represents each cluster to be merged by a constant number

of well scattered points that capture the shape and extent of the cluster (Abstract, line 7). The

clusters with the closest pair of representative points are merged to from a new cluster (Fig 4).

This process continues until a desired number of clusters is reached (col 4, lines 40-45).

Guha does not expressly disclose the claimed "(f) otherwise when the clustered results

are not satisfactory revise the size parameter perform clustering based on the revised size

parameter and the clustered results, and proceed to step (d)". However, Guha discloses having a

partition size that is smaller that the size of main memory. Reducing C would make sure that the

data points fit in main memory (col 16, lines 50-55). Being able to fit all data points in main

memory would result in better cluster calculations. Therefore it would have been obvious to one

having ordinary skill in the art at the time that the invention was made to modify the teachings of

Guha to reduce the C parameter. One of ordinary skill in the art would have been motivated to

do this because it would enable the clustering of large databases (col 1, lines 5-10).

Referring to Claim 2:

Guha discloses the limitations as discussed in Claim 1 above. Guha further discloses the

scattered points are used as representatives of the cluster. The clusters with the closest pair of

representative points are merged to reduce the number of clusters. The process continues until a

desired number of clusters is reached (col 4, lines 35-45).

Guha does not expressly disclose the claimed "(c 1) evaluating subsets of data points in

each cluster for moving into every other cluster by using a predetermined metric; wherein the

number of data points in the subset is specified by the size parameter."

However, the teachings disclosed by Guha are essentially the same as the claimed

limitation. The c parameter disclosed by Guha has the same functionality as the claimed size

parameter (col 6, lines 54-62), they both help in determining whether the functions will datasets

will be merged.

It would have been obvious to one having ordinary skill in the art at the time that the

invention was made to modify the teachings of Guha. One of ordinary skill in the art would have

been motivated to do this because it would enable the clustering of large databases (col 1, lines 5-10).

Referring to Claim 3:

Guha discloses the limitations as discussed in Claim 2 above. Guha further discloses that the c is representative set of the cluster (col 9, lines 30-35). This set is used to calculate the distance between clusters (col 7, lines 25-37).

Guha does not expressly disclose the claimed " (c1_1) determining a geometric center of the subset of data points being evaluated for a move; (c1_2) using the geometric center of the subset of data points in the predetermined metric to generate a value."

However, the teachings disclosed by Guha are essentially the same as the claimed limitation.

It would have been obvious to one having ordinary skill in the art at the time that the invention was made to modify the teachings of Guha. One of ordinary skill in the art would have been motivated to do this because it would enable the clustering of large databases (col 1, lines 5-10).

Referring to Claim 4:

Guha discloses the limitations as discussed in Claim 3 above. Guha further discloses that a heap structure stores the distances that are calculated (col 7, lines 44-45). The method loops until only clusters remain (col 8, lines 5-10). The first element in the cluster is merged with the given cluster (col 8, lines 15-25).

Guha does not expressly disclose the claimed "(cl_ 3) determining whether the value is greater than zero; (cl_ 4) when the value is greater than zero, moving the subset of data points from a Move From cluster to a Move To cluster; (cl_5) when the value is not greater than zero, determining if there are more subsets to evaluate;(cl_6) when there are more subsets to evaluate, proceeding to step (cl); (cl_7) when there arc no more subsets to evaluate, determining whether ally point has moved; (cl_8) when a point has moved, proceeding to step (cl); and (cl_9) when no point has moved, stopping the processing."

However, the clusters as described by Guha are very similar to the claimed subsets. The functionality of the claimed limitations and that of Guha's invention are essentially the same.

It would have been obvious to one having ordinary skill in the art at the time that the invention was made to modify the teachings of Guha. One of ordinary skill in the art would have been motivated to do this because it would enable the clustering of large databases (col 1, lines 5-10).

Referring to Claim 5:

Guha discloses the limitations as discussed in Claim 4 above. Guha further discloses that that after the merge c is re-calculated (col 8, lines 10-35).

Guha does not expressly disclose the claimed "simultaneously updating the membership of at least two data points from the membership of the Move From cluster to the membership of the Move To cluster".

However, the merge operation as disclosed by Gupta is similar to the claimed move. The functionality of the claimed limitations and that of Guha's invention are essentially the same.

It would have been obvious to one having ordinary skill in the art at the time that the invention was made to modify the teachings of Guha. One of ordinary skill in the art would have been motivated to do this because it would enable the clustering of large databases (col 1, lines 5-10).

Referring to Claim 6:

Guha discloses the limitations as discussed in Claim 4 above. Guha further discloses that there is a need to compute the distance of the new cluster to all other remaining clusters. The mean is set c is also calculated for the new cluster, and all the necessary calculations and updates are carried out (col 8, lines 28-35; Fig 7).

Guha does not expressly disclose the claimed "updating the count of the Move From cluster; updating the center of the Move-From cluster; updating the count of the Move To cluster; and updating the center of the Move To cluster."

However, Guha's invention discloses limitations that are similar to claimed limitation.

It would have been obvious to one having ordinary skill in the art at the time that the invention was made to modify the teachings of Guha. One of ordinary skill in the art would have been motivated to do this because it would enable the clustering of large databases (col 1, lines 5-10).

Referring to Claim 7:

Guha discloses the limitations as discussed in Claim 1 above. Guha further discloses

reducing the partition size to be smaller that main memory, there by reducing the input size and

the (col 16, lines 50-55).

Referring to Claim 8:

Guha discloses the limitations as discussed in Claim 1 above. Guha further discloses the

metric used to determine whether or not to merge is the distance measure (col 7, lines 30-35).

This metric used the c parameter to evaluate the distance measure.

Guha does not expressly disclose the claimed "(d_1) employing a predetermined metric

for determining whether the clustered results are satisfactory; wherein the predetermined metric

includes a geometric center of the subset of points that are being evaluated for move."0

However, the limitations disclosed by Guha are essentially the same as the claimed

limitations.

It would have been obvious to one having ordinary skill in the art at the time that the

invention was made to modify the teachings of Guha. One of ordinary skill in the art would have

been motivated to do this because it would enable the clustering of large databases (col 1, lines

5-10).

Referring to Claim 9:

Guha discloses the limitations as discussed in Claim 8 above. Guha further discloses the

expression for his metric (col 7, lines 30-35).

Guha does not expressly disclose the claimed "where U is the subset of data points being evaluated for the move, |U| is the size of U that is specified by the size parameter, m_{oo} is the geometric center of U, M_1, and m_j are the centers of the clusters and n_i, and n_j, are the counts of the clusters."

However, the metric expression disclosed by Guha is essentially the same as the claimed limitation.

It would have been obvious to one having ordinary skill in the art at the time that the invention was made to modify the teachings of Guha. One of ordinary skill in the art would have been motivated to do this because it would enable the clustering of large databases (col 1, lines 5-10).

Referring to Claim 10:

Guha discloses the limitations as discussed in Claim 1 above. Guha further discloses data mining applications (col 1, lines 5-10).

Guha does not expressly disclose the claimed "utilized in one of a data mining application, customer segmentation application, document categorization application, scientific data analysis application, data compression application, vector quantization application, and image processing application."

However, the limitations disclosed by Guha are essentially the same as the claimed limitations.

It would have been obvious to one having ordinary skill in the art at the time that the invention was made to modify the teachings of Guha. One of ordinary skill in the art would have

been motivated to do this because it would enable the clustering of large databases (col 1, lines 5-10).

Prior Art

5.   The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

US 6012058 issued to Fayyad, Usama et al. Fayyad discloses on a first iteration through the FIG. 4 process the set DS (FIG. 6A) is empty. Updates to the set DS are performed at the step 130 for each cluster of the range, i=1, . . . K. For each cluster i the processor determines which singleton data elements (elements of the set RS of type SDATA), assigned to cluster i will nor change cluster membership over future data samples. These points will be used to augment the i-th element of the set DS which contains the sufficient statistics summarizing these singleton points. These points are removed from the set RS and used to update the sufficient statistics for the i-th cluster of the set DS.

Conclusion

Any inquiry concerning this communication or earlier communications from the

examiner should be directed to Monplaisir G Hamilton whose telephone number is 1703-305-

5116. The examiner can normally be reached on Monday - Friday (8:00 am - 4:30 pm).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, Kim Y Vu can be reached on 1703-305-4393. The fax phone numbers for the

organization where this application or proceeding is assigned are 1703-746-7239 for regular

communications and 1703-746-7238 for After Final communications.

Any inquiry of a general nature or relating to the status of this application or proceeding

should be directed to the receptionist whose telephone number is 1703-305-3900.

Monplaisir Hamilton
July 25, 2003

JEAN M. CORRIELUS
PRIMARY EXAMINER